Published in final edited form as:
Mol Phys. 2011 January ; 109(2): 301–313. doi:10.1080/00268976.2010.521202.

Geometrical Analysis of Cytochrome *c* Unfolding

Kristopher G. Urie[a,+], **Ekaterina Pletneva**[b], **Harry B. Gray**[a], **Jay R. Winkler**[a], and **John J. Kozak**[c]

[a]Beckman Institute, California Institute of Technology, Pasadena CA 91125-7400

[b]Department of Chemistry, Dartmouth College, Hanover, New Hampshire 03755-3541

[c]DePaul University, 243 South Wabash Avenue, Chicago, IL 60604-2301

Abstract

We have developed a geometrical model to study the unfolding of iso-1 cytochrome *c*. The model draws on the crystallographic data reported for this protein. These data were used to calculate the distance between specific residues in the folded state, and in a sequence of extended states defined by n= 3, 5, 7, 9, 11, 13, and 15 residue units. Exact calculations carried out for each of the 103 residues in the polypeptide chain demonstrate that different regions of the chain have different unfolding histories. Regions where there is a persistence of compact structures can be identified, and this geometrical characterization is fully consistent with analyses of time-resolved fluorescence energy-transfer (TrFET) data using dansyl-derivatized cysteine side-chain probes at positions 39, 50, 66, 85, and 99. Our calculations were carried out assuming that different regions of the polypeptide chain unfold synchronously. To test this assumption, we performed lattice Monte Carlo simulations to study systematically the possible importance of asynchronicity. Our calculations show that small departures from synchronous dynamics can arise if displacements of residues in the main body of the chain are much more sluggish than near-terminal residues.

1. Introduction

The hydrophobic effect drives protein folding in aqueous media [1,2]. Whereas a variety of attractive interactions, such as hydrogen bonding, salt linkages and the like, also contribute to stabilize a native structure, other factors such as repulsive interactions attributable to the space-filling character of the many residues comprising the polypeptide chain must be taken into account to have any chance of predicting the global minimum in a folding energy landscape.

Ramachandran [3] was among the first to point out that an accounting of repulsive interactions (only) leads to an effective restriction of the conformations available to a protein. Indeed, in analogy to modern theories of the liquid state, his work on the sterically-accessible regions of the phase space (the phi/psi diagram) provides a “ground-state” metric in terms of which the nuanced, native structure of a particular protein is ultimately determined by the nature and strength of the attractive forces at play. In this contribution, we explore whether and to what extent the relaxation of the steric interactions between and among the residues of a polypeptide chain leads to a preliminary understanding of the initial stages in the unfolding of a protein.

Correspondence to: John J. Kozak.
[+]present address: Field Museum of Natural History, 1400 S. Lake Shore Dr., Chicago, IL 60605-2496

Denatured proteins are heterogeneous ensembles of rapidly exchanging polypeptide conformers. Single-molecule fluorescence energy-transfer (FET) measurements highlight the existence of many different conformations in denatured proteins and during folding [4,5]. Although ensemble-averaged measurements of FET efficiency do not yield any information about the conformational heterogeneity of unfolded molecules, time-resolved FET measurements (TrFET) allow resolution of underlying distance distributions, providing information similar to that obtained from single-molecule studies [6–12].

In prior studies, we have employed TrFET measurements to elucidate both short- and long-range interactions as well as conformational heterogeneity in unfolded yeast cytochrome *c* [13]. The covalently-bound heme group in cytochrome *c* served as the energy acceptor for AEDANS fluorophores (Dns) bound to engineered cysteine residues at each of six different locations in *S. cerevisae iso*-1 cytochrome *c* (cyt): on three different helices in mutants K4C, E66C, and K99C; and in three different loops in mutants H39C, D50C, and L85C. To minimize structural perturbations, labeling sites were selected on the basis of high solvent exposure of the native residue in the folded protein.

Analysis of TrFET data for GuHCl-unfolded variants of ferricytochrome *c*. revealed six sets of ***D-A*** distance distributions ($P(r)$) that provided a detailed view of structural heterogeneity of the unfolded ensemble. The $P(r)$ distributions suggest that unfolded cytochrome *c* is not a simple random coil: extended structures coexist with more compact ones. Misligation promotes but is not the sole cause of compact structures in the unfolded ensemble: compact regions persist even when misligation is inhibited. TrFET data revealed relatively large populations of compact structures in unfolded Dns_{50}-cyt, Dns_{39}-cyt, and Dns_{66}-cyt; these structures are possibly stabilized by hydrophobic interactions between the polypeptide chain and the heme.

We have compared the experimental TrFET results to a geometrical model of cytochrome *c* unfolding that does not include any interatomic forces, but implicitly incorporates the native structure of the protein. In this sense, our model is similar in spirit to the geometrical approach taken by Ramachandran in his development of phi/psi plots for the native state [3].

2. Methods

The crystallographic structure of iso-1 cytochrome *c* has been determined and the coordinates of each atom of each residue have been reported [14]. Using these data, one can calculate the distance between the alpha carbon atoms (C_α) of adjacent residues for each of the 103 residues of the polypeptide chain. These nearest-neighbor C_α distances are virtually constant, with the average value of 3.8 Å (Figure 1a).

A plot of C_α separation between second-nearest-neighbors reveals considerable nonuniformity in $R_{i-1}-R_{i+1}$ distances (Figure 1b). Indeed, this nonuniformity becomes even more pronounced when one calculates the distances for each residue R_i: R_{i-2} to R_{i+2}, R_{i-3} to R_{i+3}, R_{i-4} to R_{i+4}, R_{i-5} to R_{i+5}, and R_{i-6} to R_{i+6} (Figure 1c–g).

The nonuniformity of C_α separations (Figure 1b–g) is a consequence of the higher-order protein structure that is determined by the sequence of amino-acid side chains in the polypeptide chain. These residue-specific differences in C_α separation reflect the geometry of the protein in its native state.

The smallest modular unit that preserves the correct angle between second-nearest-neighbor alpha carbon atoms is the $[R_{i-1}, R_i, R_{i+1}]$ triplet [3]. Using the $[R_{i-1}-R_{i+1}]$ distance as a metric, one can estimate for each *n*-residue segment of the polypeptide chain the distance between the two terminal residues in a fully-extended configuration. For example, using the

residue-specific, triplet data displayed in Figure 1b, the distance between residues R_{i-2} and R_{i+2} in an extended configuration of $n = 5$ residues [R_{i-2}, R_{i-1}, R_i, R_{i+1}, R_{i+2}] can be calculated by summing the distance between R_{i-2} and R_i, and the distance between R_i and R_{i+2}. As noted earlier, the *actual* distance between R_{i-2} and R_{i+2} in the native state can be determined from the crystallographic data (Figure 1c).

We now specify a Cartesian coordinate system in which the heme is centered at the origin and specify the radial distance to the heme from each cytochrome *c* residue in the folded protein (Figure 2). Using the Law of Cosines, the radial distance that the central residue R(i) in the above five-residue segment would have to be displaced from its position in the native state to its position in the extended state can be calculated; as a consequence, the length of each of the bars in Figure 2 will change. Carrying out this calculation for all 103 residues yields a three-dimensional, geometrical representation of a possible *first* stage of protein denaturation.

To elaborate *successive* stages in the unfolding of the iso-1 cytochrome *c* polypeptide chain, analogous calculations have been carried out for each of the 103 residues in modular segments of five, seven, nine, eleven, thirteen and fifteen residues; for each case, the displacement from the heme of the central residue R_i is determined. In this way, the (three-dimensional) unfolding of the polypeptide chain from the native state through a *sequence* of increasingly extended configurational states is geometrically characterized. We emphasize that these stages are elaborated using *only* the Law of Cosines and the crystallographic data for the protein; *no* other assumptions are involved in the development of these geometrical "snapshots". A detailed example of this geometric unfolding process is given in the Appendix.

3. Results

3.1. Cytochrome *c* unfolding

To illustrate the results, we focus on the behavior of residues 39 and 85 in the unfolding protein (Figures 3, 4). The black dot in each figure corresponds to the location of the selected residue (39 or 85) and the blue dot indicates the location of the heme. The first box (Figures 3a, 4a) is a view of the native configuration of the protein. The following five images (Figures 3b–f, 4b–f) display the evolution of the system through a series of increasingly extended configurational states, calculated as described above. The scales of the coordinate axes vary continuously throughout this series of plots in order to display more clearly the polypeptide configuration. In the final image of each figure (Figures 3g, 4g) the native and last-calculated configurations (extended state, the 13-segment case) are overlaid to illustrate the extension of the polypeptide that accompanies the simulated unfolding. The relative disposition of residues 39 and 85 relative to the heme in the native and final extended states would appear to be different. We will quantify this difference below when we consider the full complement of data for residues 39, 50, 66, 85 and 99. Here, we simply note that the configurational changes determined geometrically using the residue-specific information encoded in the triplet distances result in unfolding profiles that are qualitatively different for residues 39 and 85. The geometrical model predicts that different regions of the native protein could have different denaturation patterns. We return to this point below.

3.2. Synchronous vs. asynchronous unfolding

In our prior study of α-synuclein, some consequences of synchronous versus asynchronous motion of the polypeptide chain in influencing the efficiency of electron transfer between donor and acceptor were noted and quantified [15–17]. With the foregoing geometrical model, we found that two residues (H_{39} and L_{85}) assume different positions relative to the

heme as the protein unfolds. But, in carrying out the calculations, we assumed that the unfolding was synchronous, (*i.e.*, the stage-by-stage extension of all segments of the polypeptide chain was "lock step"). It is important to address and quantify the possible differences that might arise upon relaxing this assumption (*i.e.*, to assess the consequences of synchronous vs asynchronous unfolding of the polypeptide chain).

We have carried out lattice-based Monte Carlo simulations on various heme-R_i–R_j triads where residues *i* and *j* are members of the set used in the TrFET experiments reported earlier [18]: 39, 50, 66, 85 and 99. In our simulations, the heme is taken to be stationary, but residues *i* and *j* are free to move either synchronously or asynchronously. The native-state configuration of a given triad is taken as the initial condition (state) with the coordinates of each member of the triad determined from the crystallographic data.

To implement the Monte Carlo simulations on a Cartesian lattice, the coordinates for each residue (in Ångstroms) were scaled by 3.8 Å and the result rounded to the nearest integer. The integer [*x*,*y*,*z*] coordinates for the heme-H_{39}-E_{66} triad for example, are [−2, 4, −1], [1, 2, −1], and [−3, 2, −3], respectively. Plots of the full polypeptide chain constructed using integer coordinates adequately represent those generated with the exact coordinates.

The computational "rules" governing the synchronous (asynchronous) motion of, say, H_{39} and E_{66} are: (1) H_{39} and E_{66} cannot collide and occupy the same lattice site, nor can they "swap" positions (which also counts as a collision); (2) H_{39} is restricted to a maximum Euclidean distance of 21 integer units from the heme; and, (3) H_{39} and E_{66} are allowed a maximum separation distance of 27 integer units. The host Cartesian lattice is taken large enough to eliminate spurious boundary effects.

Each trial is carried out for *t* time steps and, to obtain good statistics, each trial is repeated *r* times (see below). During the simulation, a given heme-residue distance occurs multiple times. To characterize the temporal behavior of a given residue, we calculate its frequency distribution: the mean of the number of occurrences of a given distance over *t* time steps.

The frequency distributions of *synchronous* motion for the H_{39} and E_{66} residues after 1000, 3000, 10,000, and 100,000 time steps are shown in Figure 5. In Figure 6 we present the frequency distributions of synchronous motion for residues 39 and 50 in the triad, heme-H_{39}-D_{50} after 100,000 time steps, and the same indicator for the synchronous motion of L_{85} and K_{99} in the heme-L_{85}-K_{99} triad.

We fit the data from the Monte Carlo simulations to Gaussian distribution functions for each residue in a given triad. For example, in Figure 7 we show the Gaussian fit to the frequency distribution of residue 85 in the heme-L_{85}-K_{99} triad after 200, 400, 600, 800 and 1000 (synchronous) time steps. Similar plots can be constructed assuming various degrees of asynchronicity for each residue in each triad.

To illustrate the effects of *asynchronous* motion in a specific case, we display in Figure 8a plots of the mean of a Gaussian fit to the frequency distribution for the heme-H_{39}-L_{85} system; in Figure 8b we show the plot of the standard deviation of the Gaussian fits to the frequency distribution for the same system. We find that both the mean and the standard deviation of the Gaussian profiles increase as the system evolves, behavior similar to that documented in the experimental study [18].

3. Discussion

In Section II, we mobilized a geometrical model to display successive stages in cytochrome unfolding by tracking the residues H_{39} and L_{85} (Figures 3, 4). In Figure 9 we complement

these results, summarizing for each of five residues [18] (H_{39}, D_{50}, E_{66}, L_{85}, and K_{99}) three-dimensional representations showing: (1) the positions of the residue R_i and heme in the native state; (2) their positions in the most extended state for which calculations were performed; and, (3) a composite figure in which the spatial configuration of the protein in its native state and in the most extended state can be compared.

While instructive in giving a three-dimensional perspective of the unfolding of cytochrome, Figures 3, 4, and 9 appear to distort certain regions of the polypeptide chain. These distortions are an artifact, a consequence of the particular orientation of the grid used in the computer graphics display. We can, however, develop a more quantitative measure to document the actual configurational changes in different regions of the polypeptide chain as the protein unfolds. We use the data generated in our geometrical study to calculate the distance of a given residue R_i from the heme in the native and each of the extended states, then define the ratio of the two as $\Delta_i = R_{i\ extended}/R_{i,native}$. Values of Δ_i greater than unity indicate that the residue has moved "outward" from the heme as the protein unfolds. Displayed in Figure 10 for each of the 5-, 7-, 9-, 11-, and 13-residue configurational units specified earlier are the Δ_i values for each residue R_i.

The data represented in Figure 10 can be used to construct a region-specific, "integrated" measure of the unfolding behavior of different regions of the polypeptide chain. Determining the overall average value of $<\Delta_i>$ for the residues (#40–#57), for the residues (#64–#68), for the residues (#72–#79), and for the residues (#94–#99) in the first stage of unfolding (the 5-residue metric) yields, respectively, 1.362, 1.679, 1.543, and 1.782. Analysis of the TrFET data led to the conclusion that the relative populations of *compact* structures in the unfolded protein vary from one region to another [18]. Dns_{50}-cyt was found to have the largest fraction of compact structures, followed by Dns_{39}-cyt and Dns_{66}-cyt. For Dns_{85}-cyt and Dns_{99}-cyt, only ~10% of the unfolded ensemble corresponds to compact structures. If one associates values of Δ_i in the vicinity of unity to reflect the persistence of compact structures, the quantitative values of $<\Delta_i>$ noted above are entirely consistent with these conclusions.

The color coding of each of the profiles displayed in Figure 10 also relates to Englander foldon units [19]. Note that H_{39} and D_{50} reside in the Englander yellow/orange foldon region, E_{66} in the green foldon region, and K_{99} in the blue foldon region. Only the red foldon region is absent a specific residue "marker" in the TrFET study [18]. We observe that there is consistency between the foldon [19] and TrFET experimental results, and the quantitative results obtained in the present study.

Returning to Figures 3 and 4, the panels in each of these figures track the change in the positions of H_{39} and L_{85}, respectively, as the polypeptide unfolds. All regions of the polypeptide chain were assumed to unfold *synchronously*. To assess the consequences of relaxing this assumption, we presented in Figure 8a plots of the mean of a Gaussian fit to the frequency distribution for L_{85} as a function of the number of time steps for various degrees of *asynchronicity*. For comparison, the synchronous case for the heme-H_{39}-L_{85} triad was included and denoted (heme-H_{39}-L_{85}). For the asynchronous cases studied, the notation (async: 0-10-100) means that the heme is stationary, that there is a 10 % probability that H_{39} will move in the next time step, and a probability of 100 % that L_{85} will move in the next time step.

As is evident from Fig. 8a, differences in synchronous *vs.* asynchronous unfolding are essentially negligible on short time scales, but gradually become more pronounced with increasing evolution of the system. The most significant departures from synchronicity arise when there is a *less than* 100% probability that H_{39} will move in the next time step. Notice

that the mean systematically increases as the probability of residue 39 moving in the next time step decreases; see the curves corresponding to (0-50-100), (0-10-100), and (0-0-100). Conversely, if we assign a 100 % probability that the mid-chain residue moves in the next time step, there is essentially *no* difference in the statistics generated among (0-100-50), (0-100-10), and the purely synchronous case.

These results suggest that if a residue in the middle of the polypeptide chain undergoes with 100% probability a displacement in the next time step, more sluggish motion of near-terminal residues does not result in a significant departure from the dynamics of totally synchronous motion. More significant departures from synchronous dynamics can occur when residues in the middle of the polypeptide chain do not move in every time step, even though the motion of near-terminal residues can be quite fluid. In terms of the study presented here, we conclude that to a good approximation the stage-by-stage unfolding of different regions of the polypeptide chain can be modeled as synchronous processes.

4. Concluding Remarks

We have developed a geometrical model to study the early stages in the unfolding of cytochrome *c*. The starting point in our analysis is the crystallographic structure. In thermodynamic terms, this structure necessarily reports on the entire free energy landscape of the protein in its *native* state; all inter-residue contacts in the *folded* protein are encoded (exactly).

Motivated by our experimental studies on the *early* stages in the *unfolding* of cytochrome *c* [13,18], our geometrical model “tracks” the consequences of perturbing the (already optimized) structure of the native state of cytochrome c. This perturbation is incremental, focusing on the systematic relaxation of steric constrains (only). In particular, keeping intact the geometry of each triad of residues, we study (via the direct application of classical Euclidean geometry) the steric consequences of chain unfolding. A series of 5-residue, 7-residue, …, 15-residue “snapshots” are taken of the change in the local environment of each residue as the polypeptide chain unfolds.

The approach taken stands in contrast to applications of the Go model as implemented via molecular dynamics simulations and predicated on a set of defining assumptions, *e.g.*, that nonnative interactions do not contribute to the shape of the global free energy surface [20–23]. The principal objective of these studies is to understand the problem of protein *folding*. Our objective here was quite the reverse, to understand the *early* stages in protein *unfolding*.

Given that our approach is specifically tailored to describe early stages in the evolution from the native state, we do not claim that our geometrical model can be extended to characterize an eventual globular or random-coil state, or cast light on the probability of observing inter-residue contacts in the *unfolded* state. However, the advantage of our approach is that, once the problem has been formulated and the issue of synchronicity addressed, the geometrical consequences of perturbing the *native* state are elaborated with *no* further approximations, either mathematical or computational.

Acknowledgments

Research at Caltech is supported by the National Institutes of Health (DK019038, GM068461), the National Science Foundation (CHE-0802907), and the Arnold and Mabel Beckman Foundation.

References

1. Kauzmann W. Some Factors in the Interpretation of Protein Denaturation. Adv. Protein Chem 1959;14:1. [PubMed: 14404936]
2. Tanford, C. The hydrophobic effect: formation of micelles and biological membranes. 2nd ed.. Malabar, FL: Krieger Publishing; 1991.
3. Ramachandran GN, Sasisekharan V. Conformation of polypeptides and proteins. Adv Protein Chem 1968;23:283. [PubMed: 4882249]
4. Schuler B, Lipman EA, Eaton WA. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. Nature 2002;419:743. [PubMed: 12384704]
5. Xie Z, Srividya N, Sosnick TR, Pan T, Scherer NF. Single-molecule studies highlight conformational heterogeneity in the early folding steps of a large ribozyme. Proc. Natl. Acad. Sci. USA 2004;101:534. [PubMed: 14704266]
6. Haas E, Katchalski-Katzir E, Steinberg IZ. Effect of the orientation of donor and acceptor on the probability of energy transfer involving electronic transitions of mixed polarization. Biochemistry 1978;17:5064. [PubMed: 718874]
7. Amir D, Haas E. Determination of intramolecular distance distributions in a globular protein by nonradiative excitation energy transfer measurements. Biopolymers 1986;25:235. [PubMed: 2420384]
8. Fung BK, Stryer L. Surface density determination in membranes by fluorescence energy transfer. Biochemistry 1978;17:5241. [PubMed: 728398]
9. Lakowicz JR, Gryczynski I, Cheung HC, Wang C-K, Johnson ML. Distance distributions in native and random-coil troponin I from frequency-domain measurements of fluorescence energy transfer. Biopolymers 1988;27:821. [PubMed: 3382720]
10. Navon A, Ittah V, Landsman P, Scheraga HA, Haas E. Distribution of intramolecular distances in the reduced and denatured states of bovine pancreatic ribonucleases A. Folding initiation structures in the C-terminal portions of the reduced protein. Biochemistry 2001;40:105. [PubMed: 11141061]
11. Lyubovitsky JG, Gray HB, Winkler JR. Mapping the Cytochrome *c* Folding Landscape. J. Am. Chem. Soc 2002;124:5481. [PubMed: 11996590]
12. Lyubovitsky JG, Gray HB, Winkler JR. Conformations of unfolded S-cerevisiae cytochrome c probed by fluorescence energy transfer kinetics. Isr. J. Chem 2004;44:263.
13. Winkler JR. Cytochrome c folding dynamics. Curr. Opin. Chem. Biol 2004;8:169. [PubMed: 15062778]
14. Louie GV, Brayer GD. High Resolution Refinement of Yeast Iso-1-cytochrome *c* and Comparisons with Other Eucaryotic Cytochromes *c*. J. Mol. Biol 1990;214:527. [PubMed: 2166169]
15. Lee JC, Gray HB, Winkler JR. Tertiary Contact Formation in α-Synuclein Probed by Electron Transfer. J. Am. Chem. Soc 2005;127:16388. [PubMed: 16305213]
16. Lee JC, Lai BT, Kozak JJ, Gray HB, Winkler JR. α-Synuclein tertiary contact dynamics. J. Phys. Chem. B 2007;111:2107. [PubMed: 17279794]
17. Urie KG, Angulo D, Lee JC, Kozak JJ, Gray HB, Winkler JR. Synchronous vs Asynchronous Chain Motion in alpha-Synuclein Contact Dynamics. J. Phys. Chem. B 2009;113:522. [PubMed: 19099437]
18. Pletneva EV, Gray HB, Winkler JR. Many Faces of the Unfolded State: Conformational Heterogeneity in Denatured Yeast Cytochrome *c*. J. Mol. Biol 2005;345:855. [PubMed: 15588831]
19. Maity H, Maity M, Englander SW. How cytochrome c folds, and why: Submolecular foldon units and their stepwise sequential stabilization. J. Mol. Biol 2004;343:223. [PubMed: 15381432]
20. Onuchic JN, Wolynes PG. Theory of protein folding. Curr. Opin. Struct. Biol 2004;14:70. [PubMed: 15102452]
21. Takada S. G(o)over-bar-ing for the prediction of protein folding mechanisms. Proc. Natl. Acad. Sci. USA 1999;96:11698. [PubMed: 10518512]
22. Ding F, Jha RK, Dokholyan NV. Scaling behavior and structure of denatured proteins. Structure 2005;13:1047. [PubMed: 16004876]

23. Voelz VA, Singh VR, Wedemeyer WJ, Lapidus LJ, Pande VS. Unfolded-State Dynamics and Structure of Protein L Characterized by Simulation and Experiment. J. Am. Chem. Soc 132:4702. [PubMed: 20218718]

Appendix

Appendix

The details of the geometric model used to describe cytochrome *c* unfolding are illustrated in the following specific example. For definiteness, we focus on the seven residue segment, from residue R_{01} to residue R_{07}; the "midpoint" residue is then residue R_{04}.

Figure A1 is a schematic representation of the problem; the symbols in the figure are defined below.

Relative to the heme (H), the distances to the residue R_{01} and to the residue R_{07} ($D(H\text{-}R_{01})$ and $D(H\text{-}R_{07})$) can be calculated from the crystallographic data, using the theorem of Pythagorus:

$$D(H-R_{01})=\sqrt{(R_{01,x}-H_x)^2+\left(R_{01,y}-H_y\right)^2+(R_{01,z}-H_z)^2}=20.522$$
$$D(H-R_{07})=\sqrt{(R_{07,x}-H_x)^2+\left(R_{07,y}-H_y\right)^2+(R_{07,z}-H_z)^2}=15.082$$
$$D(R_{01}-R_{07})=\sqrt{(R_{07,x}-R_{01,x})^2+\left(R_{07,y}-R_{01,y}\right)^2+(R_{07,z}-R_{01,z})^2}=11.177$$

Here, $\{H_x, H_y, H_z\}$, $\{R_{01,x}, R_{01,y}, R_{01,z}\}$, and $\{R_{07,x}, R_{07,y}, R_{07,z}\}$ are the crystallographic coordinates.

The distance $D(R_{01}\text{-}R_{07})$ is a geometrical signature of the native structure for this segment of the polypeptide chain.

When the protein begins to unfold, the intrinsic geometric structure of this segment will change; the polypeptide chain will become more "extended." One possible configuration of this more extended state is illustrated below (Figure A2):

The triplet distance $S(R_{01}-R_{03})$ can be calculated from the crystallographic data, as can the triplet distances $S(R_{03}\text{-}R_{05})$, and $S(R_{05}\text{-}R_{07})$, *viz*.

$$S(R_{01}-R_{03})=\sqrt{(R_{03,x}-R_{01,x})^2+\left(R_{03,y}-R_{01,y}\right)^2+(R_{03,z}-R_{01,z})^2}=6.354$$
$$S(R_{03}-R_{05})=\sqrt{(R_{05,x}-R_{03,x})^2+\left(R_{05,y}-R_{03,y}\right)^2+(R_{05,z}-R_{03,z})^2}=5.413$$
$$S(R_{05}-R_{07})=\sqrt{(R_{07,x}-R_{05,x})^2+\left(R_{07,y}-R_{05,y}\right)^2+(R_{07,z}-R_{05,z})^2}=5.465$$

Let us define the sum of these three distances as T_{04}, such that

$$T_{04}=S(R_{01}-R_{03})+S(R_{03}-R_{05})+S(R_{05}-R_{07})=17.233$$

We would like to determine the distance that the "*midpoint*" of the length R_{01} to R_{07} (Figure S1) moves "outward" as the polypeptide begins to unfold from the native state to the above-defined, extended state.

The first step in the calculation is to determine the angle β (Figure A1).

In a triangle with angles *A*, *B*, and *C* and sides opposite *a*, *b*, and *c* respectively, the Law of Cosines states:

$$c^2=a^2+b^2-2ab\cos(C)$$

In Figure A1, we identify

$$c=D(R_{01}-R_{07})=11.177$$
$$a=D(H-R_{01})=20.522$$
$$b=D(H-R_{07})=15.082$$
$$C=\beta$$

Solving for β gives:

$$\cos\beta=0.8460;\beta=32.222^\circ$$

Having determined β we can proceed to calculate the distance Y, the distance from the heme (*H*) to the midpoint of the base T_{04} in the Figure S1:

Since we know the distance ½T_{04} and the half-angle β/2, from the definition of the tangent, we have:

$$\tan\left(\beta/2\right)=\frac{\frac{1}{2}T_{04}}{Y}$$

which gives Y = 29.830.

If residue R_{04} were positioned *exactly* at the midpoint of the length T_{04}, the distance from R_{04} to the heme would be extended by a factor s_{04}, where

$$s_{04}=\frac{Y}{D(H-R_{04})}=1.515$$

Of course, the residue R_{04} is not (necessarily) positioned at the exact midpoint of T_{04}. But, there are two calculations we can do to explore this stage of unfolding of the seven-residue segment.

The first calculation proceeds as follows. The distance of residue R_{04} from the heme in the folded protein is given by:

$$D(H-R_{04})=\sqrt{(R_{04,x}-H_x)^2+\left(R_{04,y}-H_y\right)^2+(R_{04,z}-H_z)^2}=19.690$$

In the stage of unfolding being studied, the distance $D(H\text{-}R_{04})$ will change. Let us call this new distance $D^*(H\text{-}R_{04})$, defined formally by:

$$D*(H - R_{04}) = \sqrt{\left(R^*_{04,x} - H_x\right)^2 + \left(R^*_{04,y} - H_y\right)^2 + \left(R^*_{04,z} - H_z\right)^2}$$

If we assume uniform scaling in each of the three Cartesian dimensions, then $D^*(H\text{-}R_{04}) = s_{04} \times D(H\text{-}R_{04})$, and we obtain *via* straightforward algebra:

$$R^*_{04,x} = s_{04}R_{04,x} + (1 - s_{04})H_x = -9.202$$
$$R^*_{04,y} = s_{04}R_{04,y} + (1 - s_{04})H_y = 31.815$$
$$R^*_{04,z} = s_{04}R_{04,z} + (1 - s_{04})H_z = -27.538$$

The point $\{R^*_{04,x}, R^*_{04,y}, R^*_{04,z} = -9.202, 31.815, -27.537\}$ can be plotted on a three-dimensional grid.

Repeating the above calculation for *each* of the *i* residues of the polypeptide chain, one generates a series of points $\{R^*_{i,x}, R^*_{i,y}, R^*_{i,z}\}$. The result are displayed in manuscript Figures 3, 4, and 9) for the *seven*-residue segment case.

Alternatively, one can use explicitly the three-dimensional coordinates $\{R^*_{04,x}, R^*_{04,y}, R^*_{04,z}\}$, to calculate the distances of residue R_{04} from the center of the original crystallographic reference frame in the folded and extended protein ($d(R_{04})$, $d^*(R_{04})$):

$$d(R_{04}) = \sqrt{\left(R_{04,x}\right)^2 + \left(R_{04,y}\right)^2 + \left(R_{04,z}\right)^2}$$
$$d^*(R_{04}) = \sqrt{\left(R^*_{04,x}\right)^2 + \left(R^*_{04,y}\right)^2 + \left(R^*_{04,z}\right)^2}$$

and define the ratio $\Delta_{04} = d^*(R_{04})/d(R_{04})$.

Repeating this calculation for each residue *i*, and plotting the results as a function of the residue number i, yields the plot in Figure 10 for the seven-residue segment case.

To follow the stage-wise unfolding of the polypeptide, the above program of calculation is repeated for the n=5, 7, 9, 11, and 13 segment cases.

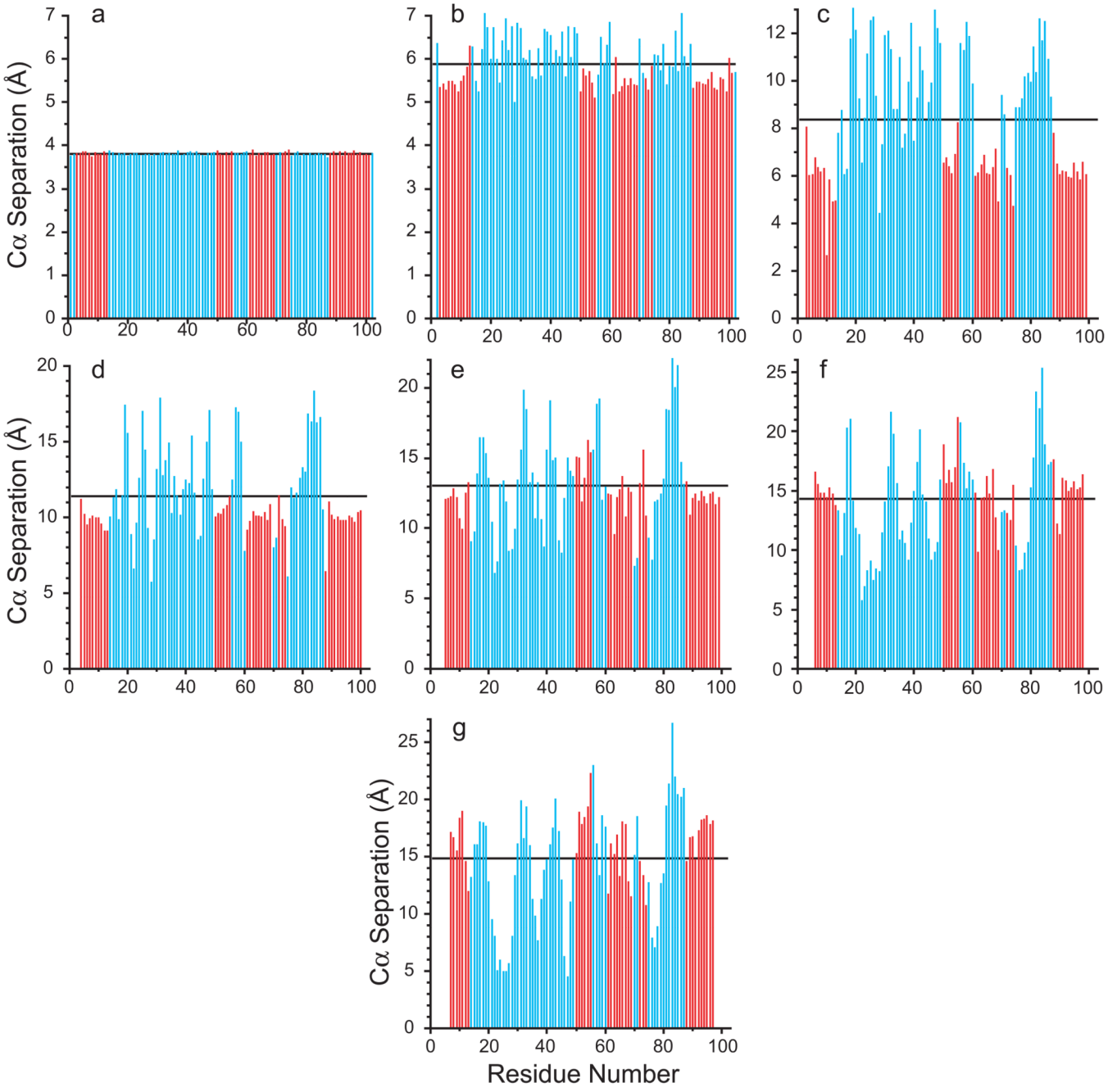

Figure 1.

Distances between n^{th}-nearest neighbors in iso-1 cytochrome *c*. Hydrophobic and hydrophilic residues are in blue and red, respectively. The horizontal black line in each plot is the average value of the distances calculated for all residues considered. (a) Nearest neighbors; (b) R_{i-1} to R_{i+1}; (c) R_{i-2} to R_{i+2}; (d) R_{i-3} to R_{i+3}; (e) R_{i-4} to R_{i+4}; (f) R_{i-5} to R_{i+5}; (g) R_{i-6} to R_{i+6}.

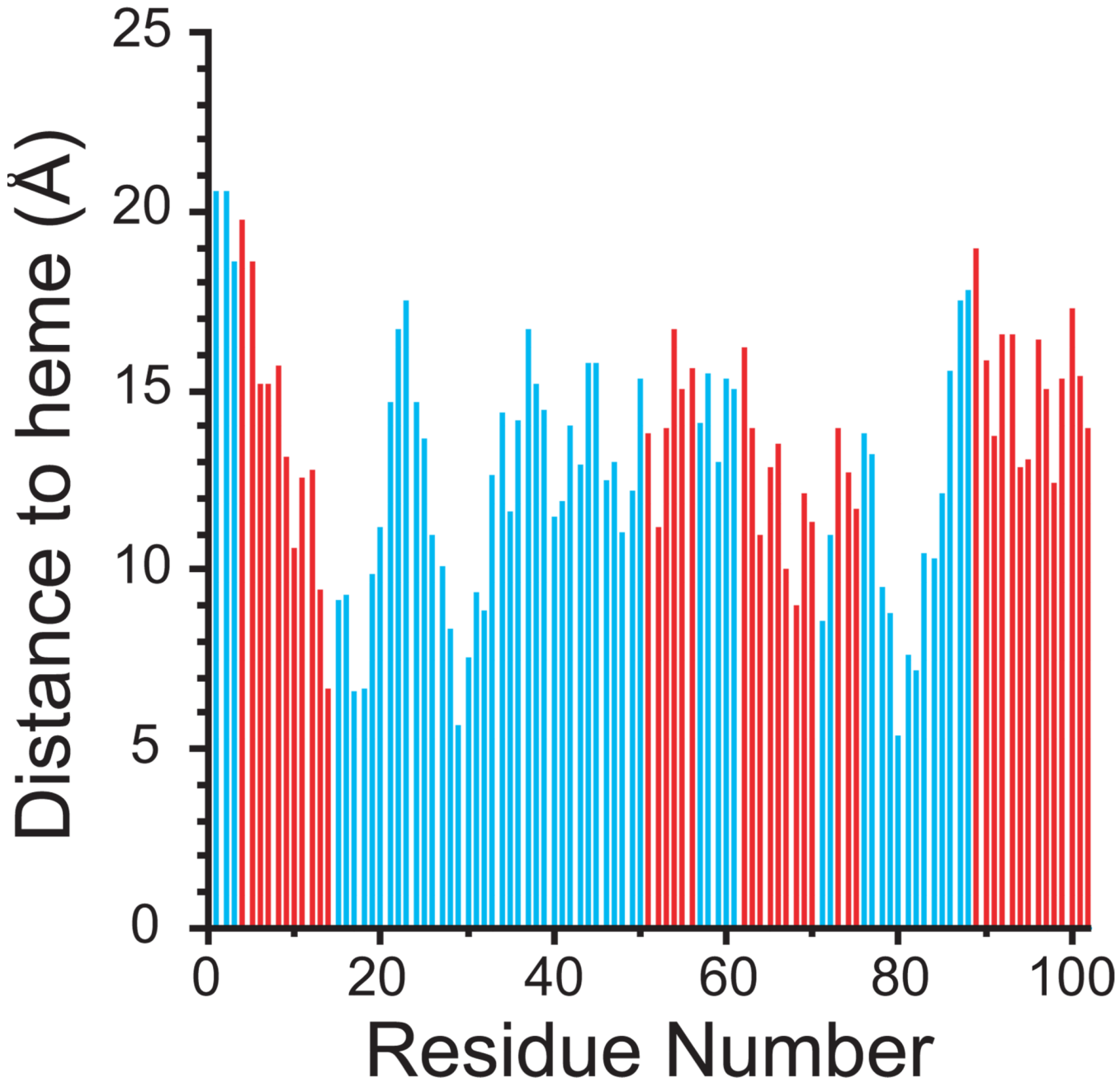

Figure 2.
Radial distance from the heme (in Å) of each residue R_i in folded iso-1 cytochrome *c* (hydrophobic and hydrophilic residues are in blue and red, respectively).

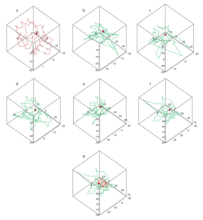

Figure 3.
Unfolding sequence in the vicinity of H_{39}; the black dot represents the location of the H_{39} and the red dot gives the location of the heme, and the red and green lines illustrate the trace of the polypeptide backbone (see text and Appendix): (a) native state; (b) $n = 5$; (c) $n = 7$; (d) $n = 9$; (e) $n = 11$; (f) $n = 13$ extended state; (g) overlay of native and $n = 13$ extended states.

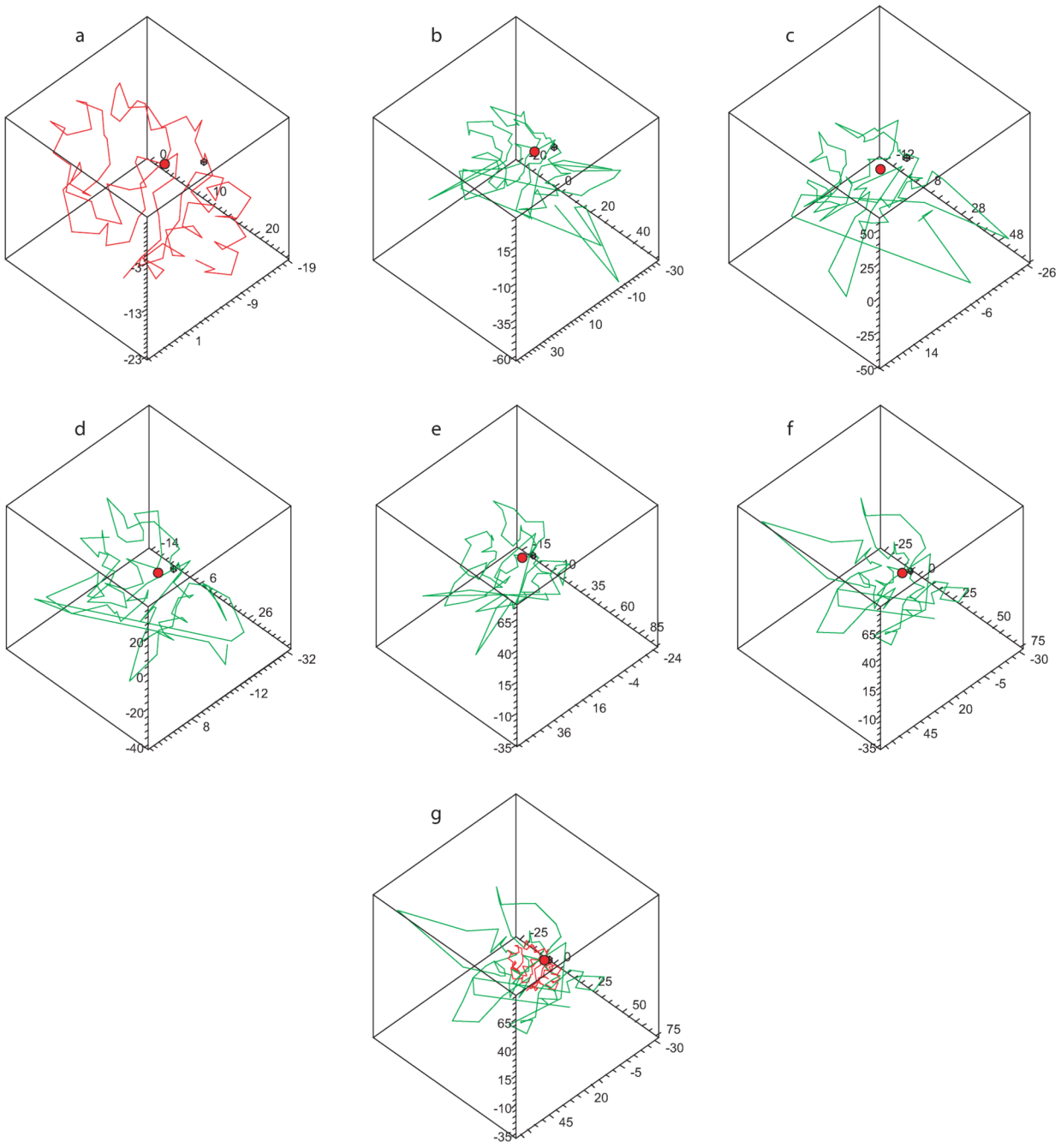

Figure 4.

Unfolding in the vicinity of L_{85}; the black dot gives the location of the residue and the red dot gives the location of the heme, and the red and green lines illustrate the trace of the polypeptide backbone (see text and Appendix): (a) native state; (b) $n = 5$; (c) $n = 7$; (d) $n = 9$; (e) $n = 11$; (f) $n = 13$ extended state; (g) overlay of native (red) and $n = 13$ extended (green) states.

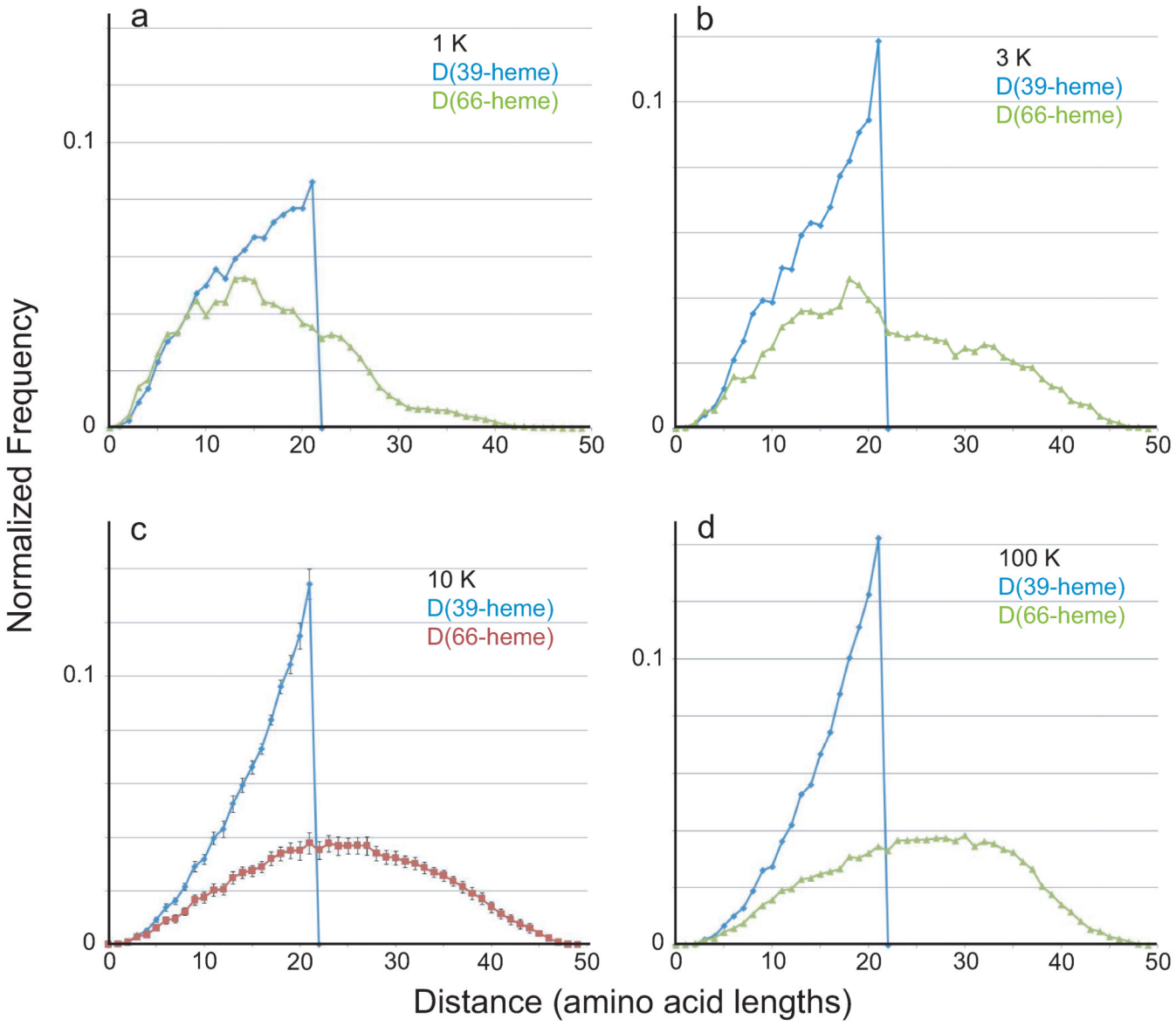

Figure 5.
Frequency distributions of heme-H_{39} and heme-E_{66} distances in the heme-H_{39}-E_{66} triad after (a) 1000, (b) 3000, (c) 10,000, and (d) 100,000 synchronous time steps.

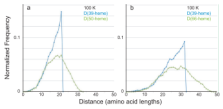

Figure 6.

Frequency distribution after 100,000 synchronous time steps for: (a) heme-H_{39} and heme-D_{50} distance in the heme-H_{39}-D_{50} triad; and (b) heme-D_{50} and heme-E_{66} distance in the heme-D_{50}-E_{66} triad.

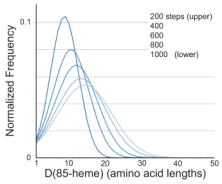

Figure 7.
Gaussian distribution fit to the frequency distribution of L_{85} in the heme-L_{85}-K_{99} triad after 200, 400, 600, 800 and 1000 times steps (upper to lower).

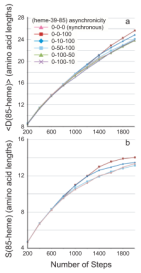

Figure 8.

The mean (a) and standard deviations (b) of Gaussian fits to the frequency distributions for L_{85} in the triad, heme-H_{39}-L_{85}, with different asynchronicity values.

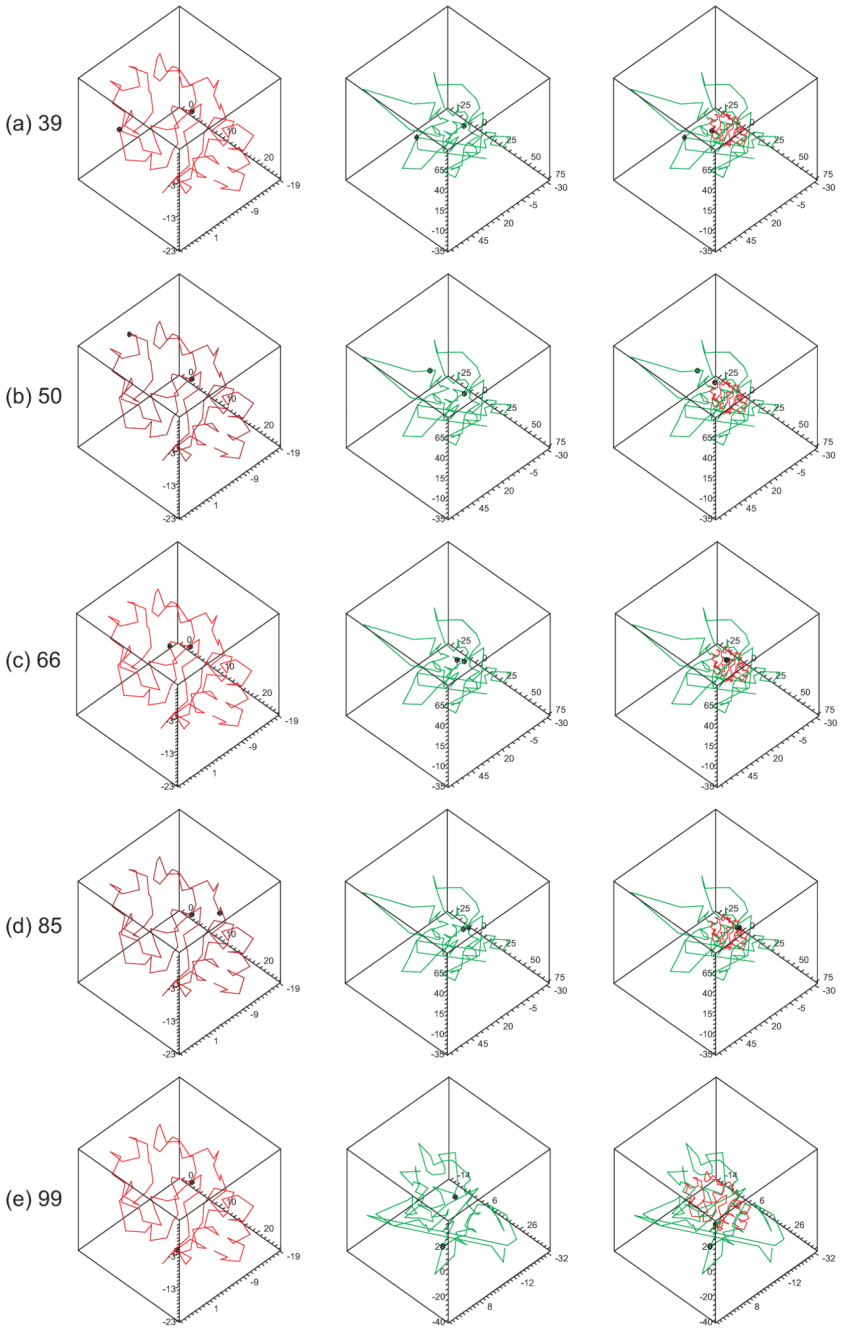

Figure 9.

Unfolding in the vicinity of residue R_i; the black dot gives the location of the residue and the blue dot gives the location of the heme: native and $n = 13$ extended-state configurations for (a) H_{39}, (b) D_{50}, (c) E_{66}, (d) L_{85}; (e) native and $n = 9$ extended-state configurations for K_{99}.

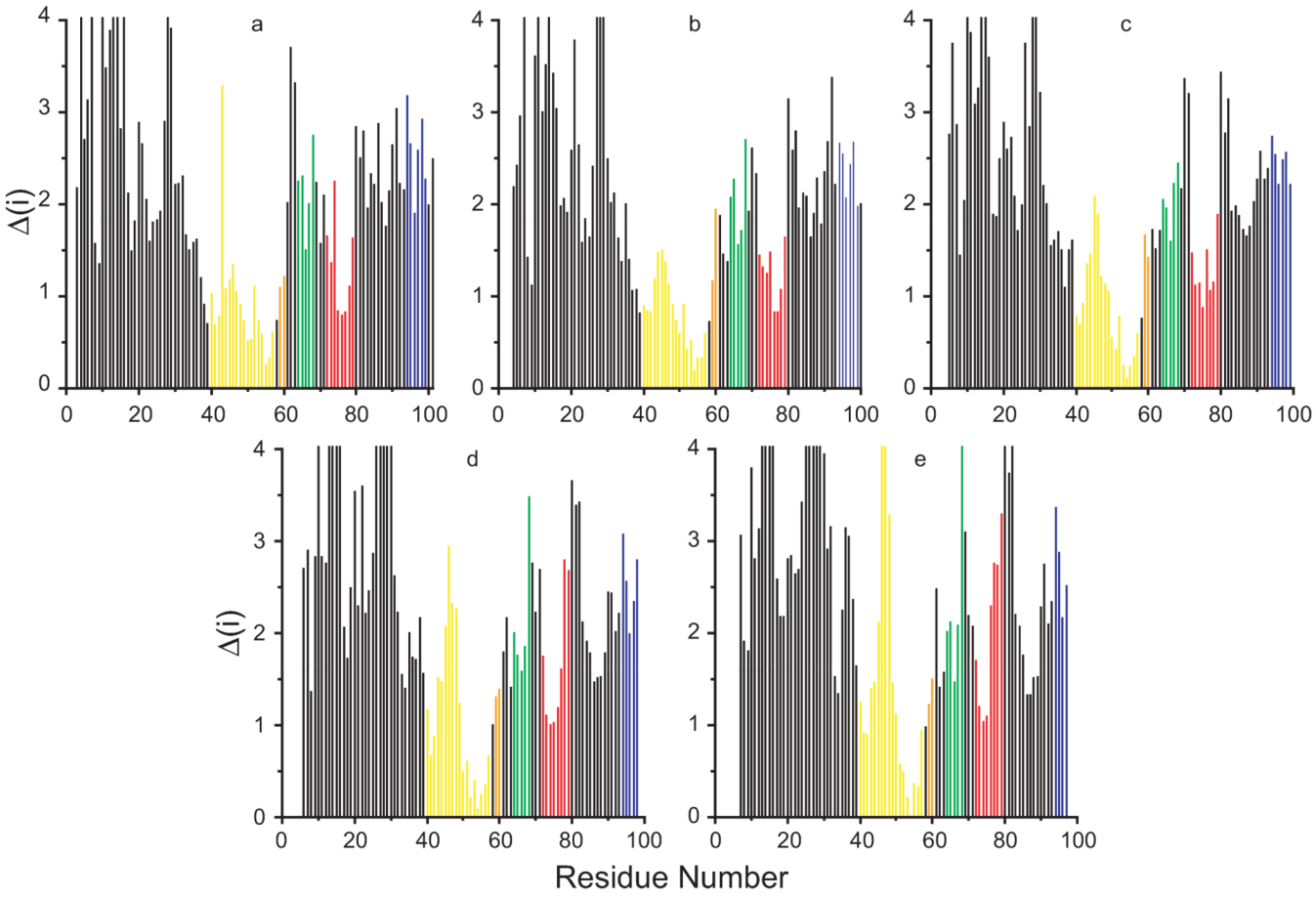

Figure 10.

Plots of the ratio Δ_i (see text and Appendix) for residues R_i: (a) configuration unit R_{i-1} to R_{i+1} for residues i = 3–101; (b) configuration unit R_{i-3} to R_{i+3} for residues i = 4–100; (c) configuration unit R_{i-4} to R_{i+4} for residues i = 5–99; (d) configuration unit R_{i-5} to R_{i+5} for residues i = 6–98; (e) configuration unit R_{i-6} to R_{i+6} for residues i = 7–97. Residue coloring follows the Englander foldon model [19].

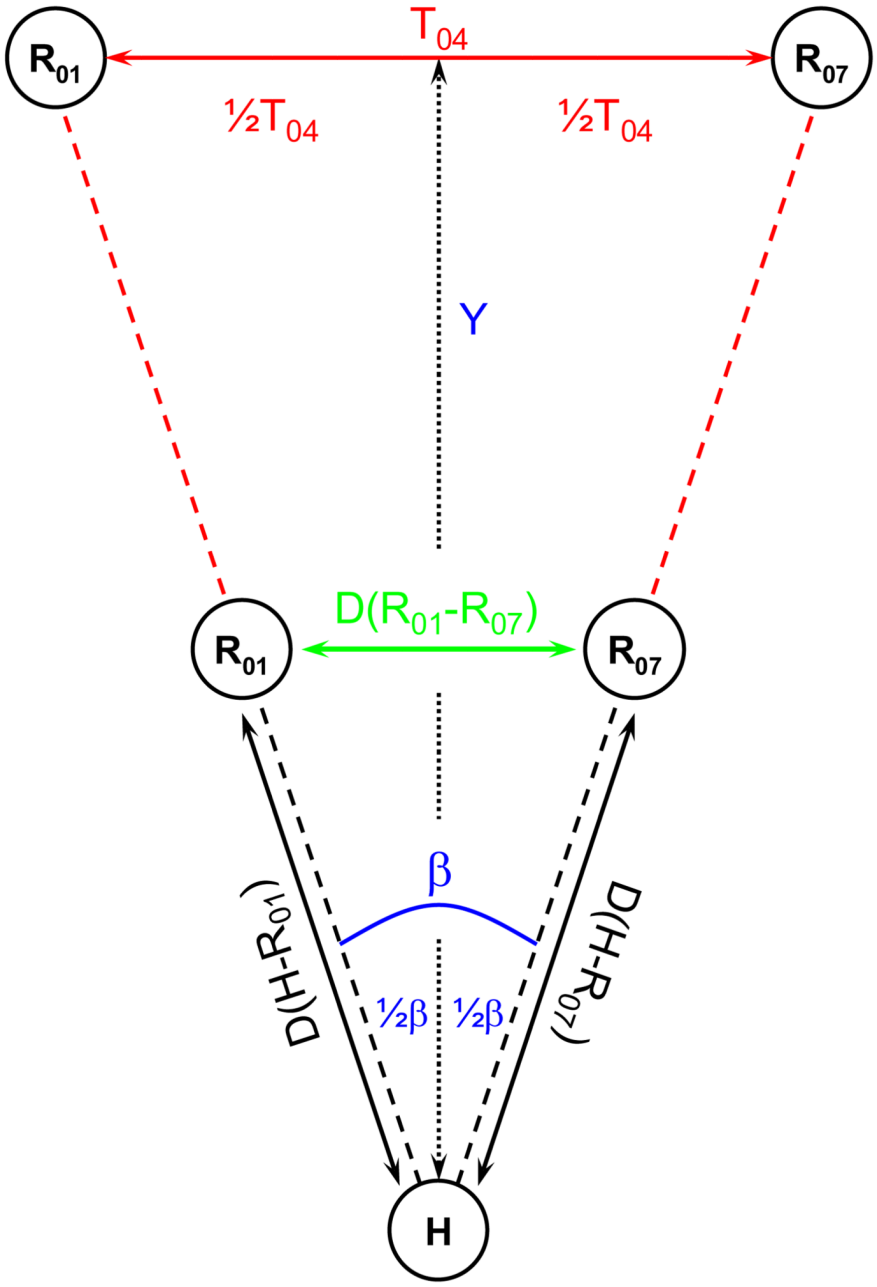

Figure A1.

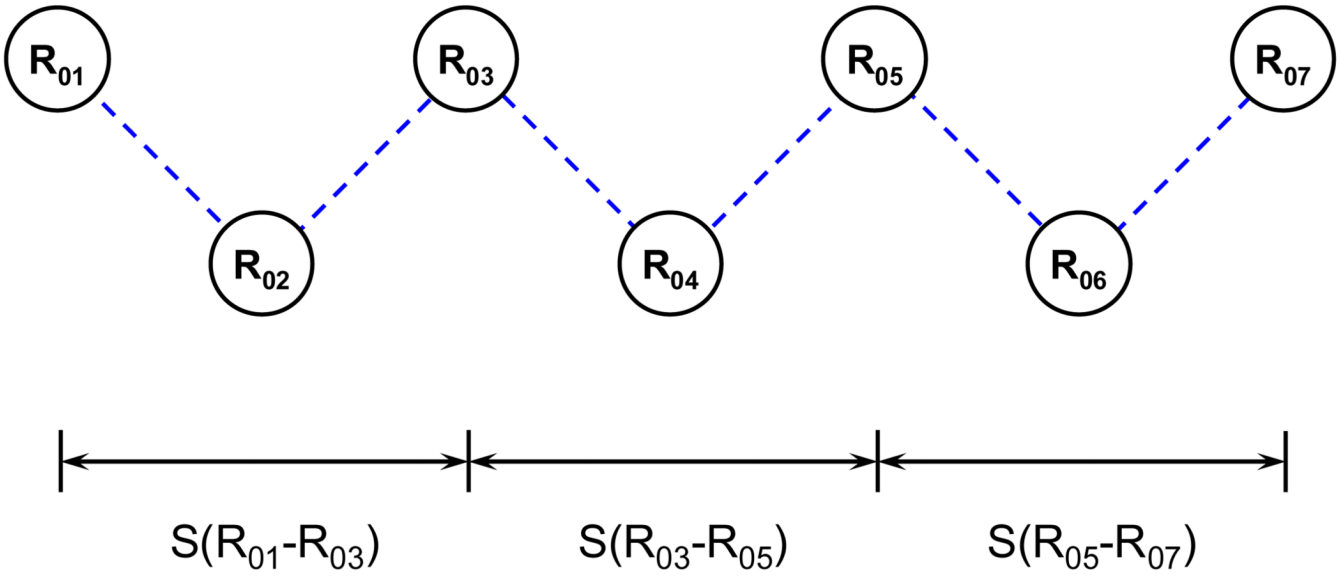

Figure A2.